

AD No. 34630

ASTIA FILE COPY

Contract N8onr - 66213 with the Office of Naval Research  
Project NR153-148: Performance Examinations for the Training  
and Selection of Scientific Personnel

Paper-pencil Analogs of Laboratory

Performance Tests

Haym Kruglak

Department of Physics  
University of Minnesota

Technical Report No. 9

June, 1954

THIS REPORT HAS BEEN DELIMITED  
AND CLEARED FOR PUBLIC RELEASE  
UNDER DOD DIRECTIVE 5200.20 AND  
NO RESTRICTIONS ARE IMPOSED UPON  
ITS USE AND DISCLOSURE.

DISTRIBUTION STATEMENT A

APPROVED FOR PUBLIC RELEASE;  
DISTRIBUTION UNLIMITED.

## The Measurement of Laboratory Achievement\*

---

\* Under contract N8 onr-66213 with the Office of Naval Research, Project  
NR153-146

---

### Part III. Paper-pencil Analogs of Laboratory Performance Tests

Haym Kruglak

University of Minnesota, Minneapolis 14, Minnesota

#### Abstract

An attempt has been made to convert laboratory performance tests into essay and multiple choice items. Preliminary forms of the tests were administered to about 160 elementary physics students. It was found that the multiple-choice was the least difficult of the three tests; the performance-- the most difficult. The correlation coefficients between relatively complex performance items and their paper-pencil analogs were very low. The correlations between the items dealing with specific skills and parallel paper-pencil tests were low to moderate. The preliminary study supports the hypothesis that paper-pencil tests are poor substitutes for performance examinations.

#### Introduction

Few experienced physics teachers would venture to predict on the basis of conventional tests what a student will accomplish when confronted with instruments, apparatus and materials; and yet the widespread use of paper-pencil examinations is based on the assumption that ability to solve a problem on paper is highly correlated with the ability to solve the same problem in the shop or laboratory.

The results of earlier studies<sup>1</sup> have indicated that: (1) performance tests appeared to measure outcomes other than those sampled by the conventional

---

<sup>1</sup> H. Kruglak, Am. J. Phys. 22, 1954

---

achievement tests in physics and (2) paper-pencil tests dealing with specific skills in the laboratory were only slightly related to performance tests containing more general and difficult tasks. If it were possible to construct paper-pencil items that were highly correlated with parallel performance tasks, then the measurement of laboratory achievement would be greatly simplified.

### The Problem

The major problem of the study could be stated as follows: To what extent is the ability to solve a laboratory problem on paper related to *The ability* solve the same problem using apparatus and materials? The "same" implies parallel content, objective, method, conclusion or results. The investigation sought the answers to a number of sub-problems: (1) Is there any difference between the essay and multiple-choice tests with analogous items? (2) If two laboratory tests are taken successively, which produces the greatest practice effect on the other, the performance or paper-pencil test? (3) Are there any laboratory activities which lend themselves to evaluation by paper-pencil analogs better than other activities?

The investigation was carried out at the University of Minnesota during the Winter quarter 1953-54. Since many of the experimental conditions were far from ideal, the study should be considered as exploratory and its findings far from definitive.

### The Experimental Procedure

#### a. Test construction

It was decided to construct tests that would sample three types of laboratory achievement: a complete laboratory problem familiar to the student from his laboratory work during the quarter; an original task unfamiliar to the student but involving relatively simple instrumental manipulations; a group of specific skills and techniques. Each of the three parts was to be prepared in three forms: performance, essay and short-answer, and multiple-choice. The preliminary drafts were prepared by the writer and Mr. R. V. Stuart. The items were criticized by Profs. C. N. Wall and G. D. Preier.

In designing the items three criteria were used as guides.

1. The items should be representative of the achievement commonly expected in the sophomore physics laboratory.
2. The items should cover specific as well as general skills.
3. The paper-pencil items should be as analogous as possible to the performance tasks.

After a great deal of rewriting and many conferences between the test constructors and critics, the final draft of the tests used in the study was acceptable to all concerned. The essay and multiple-choice forms were mimeographed; location cards were made for the performance test. Photographic reproductions of the given apparatus, meter faces, wired circuits, etc. were multilithed and assembled into a booklet to accompany the paper-pencil tests. The description of the given apparatus and the statement of the problem were identical for the three forms.

The test dealt with the laboratory aspects of Electricity. Part I called for the measurement of an unknown resistance by the wheatstone bridge method. In Part II the problem was to identify concealed circuit elements inside a three-terminal box by using a voltmeter and a dry cell. Several specific skills were included in Part III: reading a multiple range meter, identifying a wired potentiometer circuit, identifying five pieces of electrical apparatus by name, symbol and function, and drawing schematic diagrams for two wired circuits. A major difference between some of the performance and paper-pencil tasks centered about the collection of experimental data: in the laboratory the student was pretty much on his own; for the essay and multiple choice portion of the test, the data had to be presented in tabular form and the instrument readings by means of photographs. The statement of the problem in Part II of the test is reproduced below. Fig. 1 accompanied the essay and multiple choice forms.

PART II

LOCATION: 24

TIME: 13 min.

GIVEN: Box with 3 terminals labeled A, B, C., voltmeter,  
dry cell, connectors.

Between any two terminals inside the box there are the  
following possible circuit elements:

- (a) Single resistor of 10 to 50 ohms
- (b) Single low resistance cell with an emf of about  
1.5 volts
- (c) A heavy copper lead (zero resistance)
- (d) An infinite resistance (open circuit)

The resistance of the voltmeter is about 200 ohms.

PROBLEM: (1) Which of the terminals are connected by  
concealed circuit elements?  
(2) Identify the elements (resistor, battery, etc.)  
inside box and draw a schematic diagram of the  
network.

HINT: It is possible to connect the given components in  
any manner without damaging the apparatus. For  
example, some of the possible circuits are shown below.

NOTE: To get credit you must sketch schematic diagram of  
circuit used, record the data, and state the conclusions  
based on the data.

When instructor signals, move to location 35.

---

Insert Fig. 1 approximately here

---

b. Test population

The test population consisted of 83 students in Physics 5--a course for  
premedics -- and 82 students in Physics 8, mostly Institute of Technology

majors. The subjects in Physics 5 had completed two quarters of physics with College Physics by Sears and Zemansky as a textbook. The lecture material included topics in mechanics, heat, and electricity with the exception of electronics. The Physics 8 textbook was Analytical Experimental Physics by Lemon and Ference. The students in Physics 8 had completed mechanics, heat and magnetostatics, electrostatics, electromagnetism and Ohm's Law. Both groups had performed six experiments in electricity out of Physics Laboratory Manual by Wall and Levine: Electric and Magnetic Fields, Joule's Law, Condenser Capacitance, Galvanometer Sensitivity, Wheatstone Bridge, and Potentiometer. The students worked in the laboratory in pairs and submitted a weekly written report which was graded by the laboratory instructor.

#### c. Test administration

The experimental tests were administered during the last laboratory period of the winter quarter. On the day of the test the students reported to a classroom where the laboratory instructor divided the section at random into two groups. One group followed the instructor to the laboratory and took the performance test during the first half of the period; the other group remained in the classroom and was administered the paper-pencil tests. During the second half of the period the testing procedure was reversed. Thus all students were required to take the performance test; one half were given the essay form and the remaining half the multiple-choice test. The administration time for each test was 52 minutes, with 26 minutes allotted for the wheatstone bridge problem and 13 minutes each for the other two parts. The allotted time appeared to be ample for the paper-pencil tests; the students were somewhat rushed in Part III of the performance examination. The order of problem presentation was also randomized in all the three forms of the test.

#### d. Test scoring

A detailed scoring key was worked out jointly by Prof. Wall, the writer, the laboratory supervisor, and the five laboratory assistants in the two

courses. The teaching assistants had themselves taken the paper-pencil and performance tests and were thoroughly familiar with the problems and the apparatus. Each assistant graded one part or section of the performance test as well as its essay analog for all the students in both courses. The weight distribution for the three parts were: 40 points for the Wheatstone bridge; 30 points for the black box; 30 points for the miscellaneous skills. The usual correction for guessing was applied to the multiple-choice scores. Since all the items on the multiple choice had no partial credit allowed, only the total scores on each part of the three tests were comparable.

### Analysis of Data

#### a. Comparison of samples

In accordance with the testing procedure there were four samples in each of the two course populations;  $E_1$ -essay first, performance second;  $E_2$ -essay second, performance first;  $M_1$ -multiple-choice first, performance second;  $M_2$ -multiple-choice second, performance first. The analysis of variance was applied to the scores on the performance test and final examination for the preceding quarter in each of the two courses. No statistically significant differences were found between the variances and the means of each four samples and therefore they were adjudged to be random samples from the same normal population as far as ability in physics was concerned. Each sample in Physics 5 was compared with the corresponding sample in Physics 8 on the experimental laboratory tests. No significant differences between the variances and the means were found on the performance test. Of the four comparisons of the paper-pencil tests only one was significant. The scores of all the subjects in Physics 5 on the performance test were compared with the scores of all the Physics 8 students on this test. No statistically significant differences were found between the variances or the means of the two groups at the 1% level. Consequently, it was reasonable to assume that the Physics 5 and 8 groups were of comparable ability and could be pooled so as to increase the size of the samples.

b. Correlations between the test parts

The correlations between the three parts of the tests are shown in Table 1.

---

Insert Table 1 approximately here

---

Most of the correlation coefficients are not significantly different from zero; those that are have very low values. From the magnitude of the correlations it was concluded that the three parts of the tests were essentially independent of each other as had been the intent of the test constructors.

c. Comparison of the total scores on the three test forms

The standard deviations, means, their respective differences, and the correlations between the performance and paper-pencil test scores are reproduced in Table 2.

---

Insert Table 2 approximately here

---

The differences in the variabilities and averages for a given sample could be accounted for by the order of test administration, by the difference in the nature of the two tests, or by both factors. It is reasonable to assume that there are no measurable differences between the four samples on any of the three tests; thus, group  $E_2$  had an average of 48.5 and a standard deviation of 14.9 on the performance test, while group  $M_2$  had an average of 47.5 and a standard deviation of 12.5 on the same test; the differences between the two respective measures were not significant.

A comparison of the averages of the three groups to whom the three forms were administered first shows a hierarchy of difficulty. The averages for the multiple choice, essay and performance were 65.6, 54.9 and 46.5 respectively. The differences between multiple-choice and each of the other two means were significant. The difference between the essay and performance means was not significant. In addition, there was a significant difference between the

standard deviations of the multiple-choice and performance tests. These two forms can hardly be considered equivalent.

The data also indicate that the practice effects are much more pronounced in the case of the performance tests, i.e., taking the essay test first produces a greater difference in the mean of the performance test than the effect of the reverse order of presentation. The effect of the multiple choice is more pronounced than that of the essay. A likely interpretation is that the essay and multiple choice forms contain possible suggestions for the answers; the practice effect due to the performance experience is much smaller with respect to the two paper-pencil forms and is almost the same for both. The magnitudes of the correlation coefficients support the above observations; the correlations are higher for the groups who took the paper-pencil tests first. However, the values of the coefficients substantiate the low degree of relationship between performance tasks and their paper-pencil analogs.

#### d. Comparison of part scores

The statistical summary for the three parts of the tests is shown in Table 3. The wheatstone bridge problem was the easiest on the performance test and most difficult as an essay question. This shows that the reproduction

---

Insert Table 3 approximately here

---

of a previously learned skill is apparently easier than the grasp of the underlying theory. The black box was very difficult as a practical problem and relatively easy as a set of multiple choice items. It is more than likely that the evaluation of original thinking is limited by the presence of the correct answers and the systematic analysis of the problem by a series of items. There was little difference between the outcomes on the miscellaneous skills part on the three forms. Actually, in terms of the test construction these skills could be more easily imitated on paper than the other two problems. For the purpose of reading, the photograph of a meter face is almost

as good as the meter face proper. Again the correlations are highest for this part of the test.

### Conclusions

In view of the experimental limitations of the study several trends appear to be supported by the data.

1. Performance tests and their paper-pencil analogs differ in difficulty as measured by mean scores. The multiple choice form differs significantly from either the essay or the performance test.
2. There is a gain on each test caused by previous administration of one of the other forms. The gain is greater for the performance test than for either of the other two forms.
3. The low values of the correlations between the performance test scores and their paper-pencil "equivalents" indicate that the latter are at best only crude approximations to the evaluation of ability to deal with laboratory materials and apparatus.
4. The relatively small differences between the means of the paper-pencil and performance items designed to evaluate very specific skills suggest that paper-pencil analogs in this area might be successfully constructed and evaluated. However, the best techniques for approximating the real situations must be used, i.e., photographs of apparatus, three-dimensional drawings, models, etc.
5. The multiple-choice form of a laboratory test is probably the least suitable type for evaluating originality.

### Summary

Three forms of a laboratory test in Electricity were constructed and administered to pre-medical and engineering students at the University of Minnesota. The items on the essay and multiple choice forms were made as analogous as possible to the corresponding performance tests. Extensive use was made of photographic technique to simulate the actual laboratory situations on paper. The best relationship was obtained for the part of the test dealing with specific skills. The paper-pencil tests which preceded

the performance tests resulted in higher scores on them. There was in general a low degree of relationship between the performance and other forms of the test.

Acknowledgments

Special thanks are due to the following people for their contributions to this investigation: Profs. C. N. Wall and G. Freier, Messrs. H. Hill, W. Johnson, J. Landis, R. Lowen, E. Mitchell, R. Stuart, D. Swenson, and R. Vik.

Table 1. Correlations between parts of the laboratory tests.

University of Minnesota, Physics 5 and 8, 1953-54.

Group	N	paper-pencil			performance		
		$r_{12}$	$r_{13}$	$r_{23}$	$r_{12}$	$r_{13}$	$r_{23}$
E <sub>1</sub>	37	.11	.32	.20	-.11	.46**	.37*
E <sub>2</sub>	47	.04	.08	.23	.23	.22	.28
M <sub>1</sub>	37	.37*	.38	.13	.28	.33*	.41*
M <sub>2</sub>	44	.26	0	.26	-.10	.22	.04

Table 2. Standard deviations, means, differences and correlation coefficients for three laboratory tests. University of Minnesota, Physics 5 and 8, 1953-54.

Group	N	Test Order	S. D.	Difference	Mean	Difference	r
E <sub>1</sub>	37	Essay first	14.9	.2	54.9	4.5	.54**
		Performance second	14.7		59.4		
E <sub>2</sub>	47	Performance first	14.9	2	48.5	12.5**	.30*
		Essay second	12.9		61.0		
M <sub>1</sub>	37	Multiple choice first	18.4	3.3	65.6	2.5	.63**
		Performance second	21.7		63.1		
M <sub>2</sub>	44	Performance first	12.5	3.6	47.5	24.0**	.47**
		Multiple choice second	16.1		71.5		

\*\* Significant at the 1% level; \* Significant at the 5% level

Table 3. Standard deviations, means, differences, and correlation coefficients for parts of the laboratory tests. University of Minnesota, Physics 5 and 8, 1953-54.

Group	N	Test Order	Part #	S. D.	Difference	Mean	Difference	r
E <sub>1</sub>	37	Essay first Performance second	I	9.8		20.5		
			II	8.7	1.1	30.0	9.5**	.36*
				7.6		11.9		
				7.3	.3	6.2	5.7**	.19
			III	3.9		22.6		
				5.4	1.5*	23.2	.6	.55*
E <sub>2</sub>	47	Performance first Essay second	I	10.4		29.2		
				8.8	1.6	24.6	4.6*	-.05
			II	4.6		2.1		
				6.6	2.0**	14.1	12.0**	.27
			III	5.6		17.1		
				4.4	1.2*	22.3	5.2**	.54**
M <sub>1</sub>	37	Multiple choice first Performance second	I	8.9		25.7		
				10.6	1.7	30.1	4.4*	.31*
			II	9.2		18.5		
				12.3	3.1*	11.5	7.0**	.34*
			III	6.7		21.2		
				5.9	.8	21.4	.2	.75**
M <sub>2</sub>	44	Performance first Multiple choice second	I	9.0		27.8		
				8.0	1.0	25.3	2.5	.03
			II	3.9		1.8		
				10.4	6.5**	22.4	20.6**	.07
			III	6.0		17.9		
				4.4	1.6*	23.8	5.9**	.47**

# I - wheatstone bridge; 40 points

II - black box; 30 points

III - miscellaneous skills; 30 points

\*\* Significant at the 1% level

\* Significant at the 5% level